

# Course 4 · Week 3 — Bayesian, biomarkers, survival ML

## Cheatsheet — biostats\_courses

R. Heller

### Bayesian thinking

$$\underbrace{p(\theta | y)}_{\text{posterior}} \propto \underbrace{p(y | \theta)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{prior}}$$

- Priors are part of the model — specify and defend them.
- Posterior summaries: mean, median, 95% credible interval.
- No  $p$ -values; use posterior probability of direction, Bayes factors, or LOO.

### brms / Stan

```
library(brms)
fit <- brm(y ~ x + (1 | group), data = df,
  prior = c(prior(normal(0, 1), class = "b"),
    prior(exponential(1), class = "sd")),
  chains = 4, iter = 2000, seed = 42)

summary(fit)
loo(fit)           # leave-one-out cross-validation
pp_check(fit)     # posterior predictive checks
```

Prior predictive check before fitting: simulate  $y$  from the prior and confirm the implications are plausible.

### Biomarker statistics

Question	Statistic
Does a biomarker classify?	AUC, cut-point (Youden's index)
Does it add over an existing model?	$\Delta$ AUC, NRI, IDI, decision curves
Does it move prognosis?	Calibration plus discrimination
Is a cut-off reproducible?	Bootstrap CI on the cut-off

```
pROC::roc(y, biomarker) |>
  pROC::coords("best", best.method = "youden")
```

### Survival ML

Model	R
Random survival forest	randomForestSRC::rfsrc(Surv(...) ~ ., data)
Gradient-boosted Cox	xgboost with objective = "survival:cox"
DeepSurv (conceptual)	torch with partial-likelihood loss

Evaluate with **time-dependent AUC**, **integrated Brier score**, and **IPA** (index of prediction accuracy =  $1 - \text{Brier}(\text{model}) / \text{Brier}(\text{null})$ ).

```
library(timeROC)
r <- timeROC(T = df$time, delta = df$event,
             marker = pred, cause = 1, times = c(365, 730))
```

## External validation

- Never trust the apparent performance on training data.
- Minimum: split-sample or CV on internal data.
- Better: external validation in a second cohort.
- Report calibration slope / intercept, discrimination, and decision curves.

## Decision rule for Week 3

- Rare outcome with need for uncertainty → Bayesian, not bootstrap of MLE.
- New biomarker → NRI + decision curve, not just  $\Delta$ AUC.
- Survival prediction → time-dependent Brier / IPA, not global AUC.
- Any clinical claim → external validation before publication.

## Common pitfalls

- Reporting a posterior with default flat priors on unscaled predictors.
- Selecting the Youden cut-off on the full data and using the same data to evaluate sensitivity / specificity.
- Quoting C-statistic at a single time point and calling it survival ML.
- Publishing a prediction model without TRIPOD-compliant reporting.

## Further reading

- Gelman et al., *Bayesian Data Analysis*, 3e.
- Royston & Altman, *External validation of a Cox prognostic model*.